

SMIL: Multimodal Learning with Severely Missing Modality

Year: 2021 | Citations: 334 | Authors: Mengmeng Ma, Jian Ren, Long Zhao, S. Tulyakov, Cathy Wu

Abstract

A common assumption in multimodal learning is the completeness of training data, i.e., full modalities are available in all training examples. Although there exists research endeavor in developing novel methods to tackle the incompleteness of testing data, e.g., modalities are partially missing in testing examples, few of them can handle incomplete training modalities. The problem becomes even more challenging if considering the case of severely missing, e.g., ninety percent of training examples may have incomplete modalities. For the first time in the literature, this paper formally studies multimodal learning with missing modality in terms of flexibility (missing modalities in training, testing, or both) and efficiency (most training data have incomplete modality). Technically, we propose a new method named SMIL that leverages Bayesian meta-learning in uniformly achieving both objectives. To validate our idea, we conduct a series of experiments on three popular benchmarks: MM-IMDb, CMU-MOSI, and avMNIST. The results prove the state-of-the-art performance of SMIL over existing methods and generative baselines including autoencoders and generative adversarial networks.